

DOT/FAA/AR-00/02

Office of Aviation Research
Washington, DC 20591

Project Plan : L³ 3DX™ 6000 EDS Usability and System Performance Evaluation

Brenda A. Klock

Aviation Security Research and Development Division
Federal Aviation Administration
William J. Hughes Technical Center
Atlantic City International Airport, NJ 08405

November 1999

Final Report

This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation security Research and Development Library, Atlantic city International Airport, New Jersey 08405.

This document is also available to the U.S. public through the National Technical Information Service (NTIS), Springfield, Virginia 22161.



U.S. Department of Transportation
Federal Aviation Administration

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents or use thereof. The United States Government does not endorse products or manufacturers. Trade or manufacturer's names appear herein solely because they are considered essential to the objective of this report.

This report is available at the Federal Aviation Administration William J. Hughes Technical Center's Full-Text Technical Reports Page: www.tc.faa.gov/its/act141/reportpage.html in Adobe Acrobat portable document format (PDF).

1. Report No. DOT/FAA/AR-00/02		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Project Plan: L ³ 3DX6000 EDS Usability and System Performance Evaluation				5. Report Date November 1999	
				6. Performing Organization Code AAR-510	
7. Author(s) Brenda Klock				8. Performing Organization Report No. DOT/FAA/AR-99/XX	
9. Performing Organization Name and Address Federal Aviation Administration William J. Hughes Technical Center Atlantic City International Airport, NJ 08405				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Federal Aviation Administration Associate Administrator for Civil Aviation Security, ACS-1 800 Independence Ave., S.W. Washington, DC 20591				13. Type of Report and Period Covered Final Report	
				14. Sponsoring Agency Code ACS-1	
15. Supplementary Notes Draft Report prepared by Veridian Engineering and Melissa Dixon, Ph. D. and William Maguire, Ph.D., Federal Data Corporation					
16. Abstract This document describes the project plan to evaluate usability and operator performance on the L³ Communications Corporation Explosive Detection System (EDS) eXaminer 3DX™ 6000. The 3DX™ 6000 uses Computed Tomography technology to produce simulated volumetric 3-D images of passenger baggage and highlight potential explosive threats. A human factors evaluation will be conducted in Phase I on the usability (i.e., graphic user interface, image quality, and information display) of the system, as well as the consistency and readability of training materials and user's manuals. In Phase II, a laboratory test with five airport security screeners and 100 test bags (30 with simulant explosives and 70 with false alarm items) will be conducted to assess system performance. The probability of detection and of false alarms, the sensitivity measure d', the response bias measure c, and the time it takes operators to make a decision will be analyzed.					
17. Key Words Computed Tomography Explosive Detection System Aviation Security Human Factors				18. Distribution Statement This report is approved for public release and is on file at the William J. Hughes Technical Center, Aviation security Research and Development Library, Atlantic city International Airport, New Jersey 08405.	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 38	
				22. Price	

EXECUTIVE SUMMARY

This Project Plan addresses the Critical Operational Issues and Criteria for usability, training, and screener alarm resolution performance established by the Federal Aviation Administration for the L³ 3DXTM 6000 Explosive Detection System. This plan was developed by the Aviation Security Human Factors Program at the William J. Hughes Technical Center, Atlantic City International Airport, New Jersey in support of the Associate Administrator for Civil Aviation Security, ACS-1. Key FAA personnel supporting this plan are all employed with the Aviation Security Research and Development Division.

ACRONYMS AND ABBREVIATIONS

BVS	Baggage Viewing Station
$\underline{c}$	Response Bias
CT	Computed Tomography
d'	Sensitivity
3DX TM 6000	eXaminer 3DX TM 6000
EDS	Explosive Detection System
FAA	Federal Aviation Administration
FAR	Federal Aviation Regulations
FK	Flesch-Kincaid Measure of Readability
GUI	Graphic User Interface
HFE	Human Factors Engineers
IAS	Image Analysis Subsystem
IED	Improvised Explosive Device
MOP	Measure of Performance
P_d	Probability of Detection
P_{fa}	Probability of False Alarm
SDT	Signal Detection Theory
2-D	Two Dimensional
3-D	Three Dimensional

TABLE OF CONTENTS

PREFACE	III
ACRONYMS AND ABBREVIATIONS	IV
1. INTRODUCTION	1
1.1 Purpose	1
1.3 Background	1
2. FUNCTIONAL REQUIREMENTS	2
2.1 Usability Considerations	2
2.1.1 Graphic User Interface	2
2.1.2 Image Quality	2
2.1.3 Information Display	2
2.2 User Training	3
2.3 Alarm Resolution	3
2.4 Maximum Operator Throughput	3
3. SYSTEM DESCRIPTION	3
4. EVALUATION OVERVIEW	4
5. CRITICAL OPERATIONAL ISSUES AND CRITERIA	5
5.1 Issue 1 – Usability	5
5.2 Issue 2 - User Training	5
5.3 Issue 3 – Alarm Resolution	6
5.4 Issue 4 – Maximum Operator Throughput	6
6. PHASE I: HFE USABILITY EVALUATION	6
6.1 Evaluation Site	6
6.2 Evaluation Materials	7
6.3 Evaluation Procedure	7
6.4 Evaluation Analysis	7
7. PHASE II: LABORATORY TEST & EVALUATION	7
7.1 Test Site	7
7.2 Participants	7
7.3 Test Personnel	7
7.3.1 Screener Training	7
7.3.2 Screener Testing	8
7.4 Test Materials	8

7.4.1 Simulant Explosive Bags	8
7.4.2 False Alarm Bags	8
7.5 Test Procedure	8
7.5.1 System To Be Tested	8
7.5.2 Informed Consent	8
7.5.3 3DX™ 6000 Training	8
7.5.4 Laboratory Test	9
7.6 Data Analysis	9
7.6.1 Alarm Resolution	9
7.6.2 Maximum Operator Throughput	10
8. TEST LIMITATIONS	10
9. PROJECT SCHEDULES	10
10. REFERENCES	11
APPENDICES	
A - Human Factors Usability Rating Scale and Checklist	
B - Consent Forms, Personal Information Forms, and Surveys	
C - Signal Detection Theory & Measures of Sensitivity	

LIST OF TABLES

1. Project Milestones	10
2. Project Schedule	10

THIS PAGE INTENTIONALLY LEFT BLANK

1. INTRODUCTION.

Code Federal Regulations section 14, part 108 requires all air carriers in the United States to provide for the safety of their passengers and property. To comply, air carriers procure equipment and train personnel to screen passengers and their carry-on baggage before they board the aircraft. The Federal Aviation Administration (FAA) ensures compliance with these and other regulations through scheduled and random inspections at airports. Maintaining a proficient screening process is critical to the mission of aviation security, both domestically and internationally.

The FAA, working with the aviation industry, encourages the development of new equipment and procedures to improve aviation security for the traveling public. X-ray systems currently in use at airports were designed to aid the screener in the detection of weapons such as guns, knives, and explosive devices like grenades. However, because of the increased sophistication of terrorists and Improvised Explosive Devices (IEDs), detection of IEDs is difficult with existing X-ray systems.

Improvements to X-ray technologies have been made to assist screeners with the detection of explosive threats. These technologies include Explosive Detection Systems (EDSs), which signal the operator when potential explosives are identified within passenger baggage. EDSs created by InVision, such as the CTX5000 and CTX5500, utilize Computed Tomography (CT) to scan passenger baggage and have been used in airports for a number of years. L³ Communications Corporation has recently introduced a similar system, the eXaminer 3DXTM 6000 (3DXTM 6000), and this system is the focus of the current project.

1.1 PURPOSE.

This plan describes the approach to a human factors evaluation of the 3DXTM 6000 user interface as well as the evaluation of training and subsequent screener performance with the system. The plan addresses the Critical Operational Issues and Criteria and Measures of Performance (MOPs) for the system to be tested.

The successful operation of the 3DXTM 6000 requires that screeners acquire a new set of skills such as the interpretation of simulated volumetric X-ray images and the ability to view multiple views of a single X-ray image. The focus of this plan is to evaluate the degree to which the 3DXTM 6000 complies with established human factors standards (e.g., usability), the effectiveness of training provided to screeners for operating the 3DXTM 6000, and the degree to which the 3DXTM 6000 increases operator effectiveness and efficiency in detecting explosive devices. The evaluation of the 3DXTM 6000 will be completed in two phases, a laboratory study of the usability of the system and a laboratory study of screener training and system operation.

1.2 BACKGROUND.

The FAA is responsible for ensuring the safety and security of air travel. Airports pose a challenge to security since they must be readily accessible to the public. To meet this challenge, the FAA has developed a security concept involving a complex system of trained personnel,

properly maintained and calibrated equipment, and appropriate procedures. This includes screening of carry-on and checked baggage as well as passengers prior to boarding an aircraft.

The threat to civil aviation security has changed in the last decade. More sophisticated explosive devices have increased airliner vulnerability to bombings. Today, explosive devices are less likely to be prefabricated, can be made from a variety of materials, and can be made to resemble innocent objects. Timing devices have been miniaturized and digitized, further compounding the difficulties of detection with current checkpoint equipment.

For these reasons, the potential for complete aircraft destruction with great loss of life and disruption of the world's air transportation system has become a great concern. This threat has increased the need for new airport security systems and operator training in these systems.

2. FUNCTIONAL REQUIREMENTS.

Automated equipment must be capable of accurately detecting IEDs without substantially decreasing the baggage throughput rate. Based upon a previous evaluation of the 3DX™ 6000 which took place at the Aviation Security Laboratory [1], the following functional requirements are identified as key factors in the evaluation of EDSs.

2.1 USABILITY CONSIDERATIONS.

The equipment must function such that an individual with a high school education can use it effectively and rapidly. The interface design should enhance rather than impede operator performance. It also should meet the requirements specified in the Usability Checklist in Appendix A.

2.1.1 Graphic User Interface.

The Graphic User Interface (GUI) must be designed so that there is logical and meaningful correspondence between the information displayed on the monitor and the use of any consoles or devices. The console must be located in a position relative to the monitor so that it is comfortable and safe to use. Operators must be able to learn how to use the interface efficiently and effectively in a short period.

2.1.2 Image Quality.

The quality of the images in terms of color, contrast, brightness, and resolution must support effective alarm resolution. Images must be clearly visible under various lighting conditions and in conjunction with the use of all equipment functions.

2.1.3 Information Display.

Information that is critical to the decision-making process must be effectively displayed. The display of useless and/or distracting information must be minimized. The images of suspected

threats must not be obstructed by the presentation of information to the operators. The use of messages, warnings, icons, and colors must be consistent and easy to understand.

2.2 USER TRAINING.

There must be effective training in the use of a system this complex. Documentation, training materials, and user manuals must be provided for instruction on using the GUI and all 3DXTM 6000 features. Any written materials for the operators must be written at or below an eighth grade reading level in order to be easily understood by people with different educational backgrounds. The classroom training must adequately prepare the operators with the knowledge, skills, and abilities necessary to manipulate and understand the images presented by the system, as well as the procedures for resolving alarms.

2.3 ALARM RESOLUTION.

Operators must be able to effectively and efficiently resolve machine alarms that identify threats as well as false alarms. Highlighting or messages that are used to indicate a suspected threat to operators must not obstruct the threat in any way.

2.4 MAXIMUM OPERATOR THROUGHPUT.

Throughput is an important consideration. The 3DXTM 6000 has a measurable throughput rate that is independent of the operator's throughput rate, because it can store images of alarm bags and continue to process new bags. However, the effective throughput of the system (i.e., machine + operator) is influenced by the amount of time the operator takes to resolve the alarms. For example, if the measurable throughput of the machine alone is 600 bags per hour, with a 10 percent alarm rate, then the operator must process 60 alarm bags per hour or one bag per minute. If the operator can only process one bag per 2 minutes, then the maximum screener throughput is only 30 bags, which in turn reduces the effective throughput rate of the system. This project will focus on the maximum operator throughput rate, which can be characterized as the number of alarm bags the operator can clear per hour. Thus, the main measure of performance is the time it takes the operator to clear each alarm bag.

3. SYSTEM DESCRIPTION.

The 3DXTM 6000, designed by L³ Communications Corporation, is the first continuously scanning EDS for checked airline baggage. The system consists of two main components, a CT scanner and an Image Analysis Subsystem (IAS). The IAS has three subcomponents, the Explosive Detection Algorithm Computer, the Baggage Viewing Station (BVS), and the Archive Computer. The BVS, where the operator views the images, consists of a monitor and a control console.

The 3DXTM 6000 acquires volume information about the densities of objects in the bag. This information is similar in type to that obtained by a medical CT scanner and that obtained by the currently deployed CTX series of scanners. The information gathered by these scanners is three-dimensional (3-D). Like the CTX, the 3DXTM 6000 software automatically identifies objects in

the bag which may be explosives. The 3DXTM 6000 acquires this information in a continuous scan rather than a series of discrete scans, but the information is largely the same.

The 3DXTM 6000 differs from the CTX series in the way that the information is presented to the operator. On one screen, the CTX presents a view which resembles an X-ray of the entire bag. On another screen it presents multiple CT views which are slices of information through the bag. Each view is a “volumetrically rendered” [2] view of the volume information from a different angle. By rotating this view, the operator can recover depth information about the volume [3, 4]. They do not perceive a 3-D view in the sense of a stereoscope or a hologram and no special glasses are worn. Rotating the view allows the operator to see the volume of information, or a part of that volume, from a continuously changing point of view.

The 3DXTM 6000 has a number of viewing options. “Full Bag” view resembles a top-down X-ray of the bag and its contents, which is similar to the CTX SP view. In “Object” view, the screen is divided into quarters. Three of these views look like X-ray views from three different angles (top/down, front/back, side/side). The fourth window presents one of two different viewing modes. In “slice mode” the 3DXTM 6000 allows the operator to view CT slices of the bag from different orientations, whereas the CTX series only allows slices in the standard orientation. In “3-D object mode” the operator sees a view of the bag from one orientation, but the operator can rotate that view continuously to view other orientations.

The 3DXTM 6000 produces a simulated 3-D image for each bag that is scanned. The system analyzes each image and identifies potential threat objects by highlighting them. There are four operating modes for viewing bag images: Alarms Only, which displays images of bags containing alarms; Alarms/Clear, which displays bags containing alarms in simulated 3-D and cleared bags in 2-D; simulated 3-D All, which displays all scanned bags in simulated 3-D; and 2-D All, which displays all scanned bags in 2-D. The system can display the images in grayscale or in color, with red representing many organic materials (a threat), green representing metallic objects, and blue representing a density consistent with an explosive.

Other system functions include the ability to work in conjunction with bar code readers to track baggage and the ability to save an image of a bag in memory for later recall.

4. EVALUATION OVERVIEW.

Human Factors Engineers (HFEs) will evaluate the 3DXTM 6000 in two phases. The first phase will consist of HFEs using a checklist (see Appendix A) to evaluate the usability of the system against accepted standards [5].

The second phase will be a test of the training and operation of the system using five screeners who will be trained by the manufacturer or their representatives. This training will be evaluated by HFEs familiar with how the system operates. Any training materials or user's manuals supplied will be similarly evaluated.

Once the screeners are trained, they will participate in a test of system performance. This will involve each screener individually using the 3DXTM 6000 under simulated operational conditions

with real bags. Explosive simulants and false alarm items will be used in order to determine each operator's effectiveness using this equipment in threat evoking detection.

Operator performance measures of the Probability of Detection (P_d), Probability of a False Alarm (P_{fa}), the sensitivity measure (d'), and the response bias measure (c) will be gathered. The time it takes the operator to make a decision will also be measured. All data will be compiled and analyzed with results to be documented in a Test and Evaluation Report.

5. CRITICAL OPERATIONAL ISSUES AND CRITERIA.

5.1 ISSUE 1 – USABILITY.

Are there any software, hardware, or procedural aspects that degrade the performance of screeners?

Criterion 1-1 Investigative in nature.

MOP 1-1-1 Deficiencies in the usability of the GUI.

MOP 1-1-2 Deficiencies in the quality of the image.

MOP 1-1-3 Deficiencies in the display of information.

MOP 1-1-4 Any other deficiencies.

In Phase I, HFEs will evaluate the 3DX™ 6000 on three main usability issues: the GUI, image quality, and information display using the Human Factors Deficiency Impact Rating Scale in conjunction with the Usability Checklist (see Appendix A). In Phase II, Each operator will complete a personal information sheet and a system usability survey (see Appendix B).

5.2 ISSUE 2 - USER TRAINING.

Does the training adequately prepare operator to use the equipment?

Criterion 2-1 Investigative in nature.

MOP 2-1-1 The Flesch-Kincaid Measure of Readability (FK) of written documentation, training materials, and user manuals.

MOP 2-1-2 Deficiencies in operator training observed during screeners' post-training performance on the 3DX™ 6000.

MOP 2-1-3 Any other deficiencies.

In Phase I, HFEs will evaluate the reading level of all written documentation, training materials, and users' manuals that will be distributed to operators. The reading level, as measured by the

FK, should not exceed 8th grade. The formula for calculating the FK is $\text{Grade Level} = (L * 0.39) + (N * 11.8) - 15.59$, where L is the average sentence length and N is the average number of syllables per word. In Phase II, HFEs will evaluate deficiencies in user training by recording the number and type of mistakes that operators make during their post-training operation of the system's functions and procedures. Each operator will also complete a survey on the usefulness of training and system usability (see Appendix B).

5.3 ISSUE 3 – ALARM RESOLUTION.

Can operators effectively and efficiently resolve alarms for threats and clear false alarms?

Criterion 3-1 Investigative in nature.

MOP 3-1-1 Operator's mean time to make a correct decision.

MOP 3-1-2 Operator's mean time to make an incorrect decision.

MOP 3-1-3 Operators' mean P_d and P_{fa} .

MOP 3-1-4 Operators' mean d' (sensitivity).

MOP 3-1-5 Operators' mean c (response bias).

MOP 3-1-6 Number and type of procedural errors in resolving alarms.

HFEs will record the type of decision an operator makes for each bag (i.e., correct rejection, miss, hit, or false alarm) and the time it takes the operator to make each decision. Each operator's P_d , P_{fa} , d' , and c will then be calculated (see Appendix C). HFEs will also record the number and type of procedural errors that operators make during the alarm resolution process.

5.4 ISSUE 4 – MAXIMUM OPERATOR THROUGHPUT.

What is the rate at which bags can be cleared?

Criterion 4-1 Investigative in nature.

MOP 4-1-1 Operator's time to clear each bag.

HFEs will record the time it takes an operator to resolve each alarm, which will allow the calculation of the number of bags that can be cleared per hour.

6. PHASE I: HFE USABILITY EVALUATION.

6.1 EVALUATION SITE.

The Aviation Security Research and Development Laboratory at the William J. Hughes Technical Center will be the location of the usability evaluation.

6.2 EVALUATION MATERIALS.

A Human Factors Deficiency Impact Rating Scale and a Usability Checklist (Appendix A) will be used to evaluate the GUI, the image quality, and the information display of L³ 3DX™ 6000. All training materials and user's manuals that L³ Communications Corporation creates for training screeners on this system will be evaluated.

6.3 EVALUATION PROCEDURE.

Two HFEs will independently evaluate the 3DX™ 6000 using the Usability Checklist to determine any usability deficiencies of the system functions and devices.

The HFEs will also compare the training materials and users' manuals to the system interface to ensure that all written documentation provided to screeners is consistent, adequate, and written at or below the 8th grade reading level.

6.4 EVALUATION ANALYSIS.

Deficiencies in the GUI, image quality, and information display, as measured by the Human Factors Deficiency Impact Rating Scale, will be analyzed. The FK for all written documentation will be calculated.

7. PHASE II: LABORATORY TEST & EVALUATION.

7.1 TEST SITE.

The Aviation Security Research and Development Laboratory at the William J. Hughes Technical Center will be the location of the test and evaluation.

7.2 PARTICIPANTS.

Five airport security screeners will participate in the test and evaluation of the 3DX™ 6000. They must be employed as an X-ray operator, 18 years of age or older, and not pregnant.

7.3 TEST PERSONNEL.

7.3.1 Screener Training.

L³ Communications Corporation, in conjunction with AMR Corporation, will create a training course for screeners who will operate the 3DX™ 6000. L³ Communications Corporation will submit this training course to the FAA for approval. Before participating in the laboratory test, the five screeners must complete the FAA-approved training course.

7.3.2 Screener Testing.

The Test Director will monitor and guide test personnel to ensure that the test protocol is being conducted properly. Data Collectors will observe and record machine and operator responses for hits and false alarms. Baggage sorters will coordinate baggage for screening to ensure that the sequence in which bags should be presented is followed.

7.4 TEST MATERIALS.

One hundred bags that will cause the 3DX™ 6000 to alarm will be assembled for the laboratory test. Thirty of the bags will contain simulant threats and 70 will contain false alarm objects.

7.4.1 Simulant Explosive Bags.

Thirty simulant explosive threats will be assembled, recorded, and randomly inserted into representative passenger checked baggage.

7.4.2 False Alarm Bags.

Seventy objects that cause false alarms will be assembled, recorded, and randomly inserted into representative passenger checked baggage.

7.5 TEST PROCEDURE.

7.5.1 System To Be Tested.

L³ Communications Corporation's eXaminer 3DX™ 6000 CT X-ray system.

7.5.2 Informed Consent.

The Test Director will describe to potential screeners an overview of the training they will receive and the experimental procedures for the test. Prior to being trained, screeners will be asked to sign an informed consent form stating that they are willing to participate in the study and to complete a personal information sheet (see Appendix B).

7.5.3 3DX™ 6000 Training.

Prior to participating in the laboratory test, the five screeners will have completed the FAA-approved training course created by L³ Communications Corporation and AMR Corporation. This training course will consist of one week of classroom training and one week of direct experience operating the machine.

7.5.4 Laboratory Test.

All test personnel will be briefed on the test purpose and procedures. After completing the training, each screener will also be briefed on the purpose and procedures of the laboratory test. The test screeners will be asked to operate the 3DXTM 6000 as if they were actually in the airport performing their jobs. They will be informed that their detection performance will be recorded.

Each screener will be presented with all 100 test bags. They will be asked to examine the contents of each bag closely to identify potential explosive objects. They will be asked to use as many of the functions as needed to confidently reject or accept an object as a threat. The screeners will be instructed to use the procedures they learned in training to resolve an alarm.

The Data Collectors will sit behind each screener with a clear view of the monitor and record the machine's analysis (i.e., hit, false alarm) and the screener's decision (i.e., hit, false alarm) for objects in each of the 100 test bags examined. They will also record the system functions that the screener uses to examine the contents of each bag, the procedures that are used to resolve an alarm, and the time it takes a screener to make a decision for each object.

Following completion of the performance test, each screener will be debriefed and interviewed. During the interview, they will be asked to complete a survey on the usability of the system and the training they received prior to the test (see Appendix B).

7.6 DATA ANALYSIS.

7.6.1 Alarm Resolution.

The Signal Detection Theory (SDT) paradigm (see Appendix C) will be applied for system analysis (i.e., machine and operator), a simulant explosive threat that is correctly identified by the machine constitutes a machine detection or hit. An explosive threat that is not correctly identified constitutes a machine miss. A machine incorrect alarm to a non-existent explosive threat constitutes a false alarm. Similarly, a machine false alarm, which the operator concurs with, is a system false alarm.

A threat that is correctly identified by the screener constitutes an operator detection or hit. A threat that is not correctly identified constitutes an operator miss. An operator identification of a non-existent threat constitutes a false alarm. If both the machine and operator correctly identify an explosive, a system hit is achieved.

All recorded operator, machine and system hits, misses, and false alarms will be analyzed according to SDT. P_d , P_{fa} , d' , and c , will therefore be calculated and statistically analyzed. The time it takes a screener to make a decision for each object or bag will also be examined. The usability and training surveys that the screeners complete will be used in conjunction with the performance results to make recommendations for improving the system.

7.6.2 Maximum Operator Throughput.

The 3DX™ 6000 has a measurable throughput rate that is independent of the operator's throughput rate, because it can store images of alarm bags and continue to process new bags. However, the effective throughput of the system (i.e., machine + operator) is influenced by the amount of time the operator takes to resolve the alarms. HFEs will record the time it takes the operator to clear each alarm bag and calculate the number of bags the operator can clear per hour.

8. TEST LIMITATIONS.

Test location is a potential limitation for alarm resolution because laboratory testing does not offer the same distracters and pressures as in the airport environment. In addition, the laboratory environment may also increase screener attention and motivation which could result in an overestimation of detection capability.

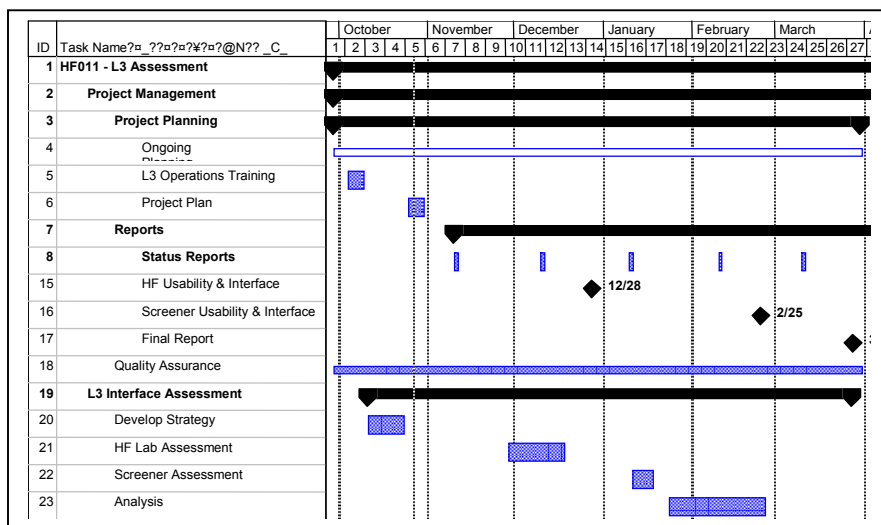
9. PROJECT SCHEDULES.

Tables 1 and 2 show the proposed milestones, resources, and overall schedule for this project.

TABLE 1. PROJECT MILESTONES

MILESTONES	DATE
Project Plan	10/29/99
Human Factors Usability and Interface Report	12/28/99
Initial Screener Interface and Usability Report	2/25/00
Final Report	3/28/00

TABLE 2. PROJECT SCHEDULE



10. REFERENCES.

1. L³ 3DX6000 Usability Assessment, (1998). Washington D.C.: Federal Aviation Administration.
2. Drebin, R.A., Carpenter, L. & Hanrahan, P., (1988). *Volume rendering*. Computer Graphics, 22 (4), 65-74.
3. Proffitt, D. R. & Kaiser, M. K. (1991). Perceiving environmental properties from motion information: Minimal conditions. In S. Ellis (Ed.). *Pictorial Communication in Virtual and Real Environments*. pp. 47-60. Bristol, PA: Francis & Taylor, Inc.
4. Wallach, H. & O'Connell, D. H. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, 45, 205-217.
5. Smith, S.L., & Mosier, J.N. (1986). *Guidelines for the design of user interface software* (Technical Report ESD-TR-86-278). L.G. Hanscom Air Force Base, MA: U.S. Air Force.

APPENDIX A
HUMAN FACTORS DEFICIENCY RATING SCALE AND USABILITY
CHECKLIST

DEFICIENCY RATING SCALE

Severity	Description
Severe	There is a high probability of operational failure, severe damage, loss of equipment, and injury to operators.
Major	There is a high probability of degraded system performance, major damage to equipment, or discomfort to operators.
Moderate	There may be no measurable impact on system performance, though there is a measurable impact upon the performance of system components or sub-systems (including the human subsystem). Operators try to compensate for, or work around, system defects.
Minimal	There is no measurable impact on the performance of system components or subsystems (including the human subsystem), although operators' negative attitudes toward features to the system may be measurable.
Negligible	The problem has a negligible impact on short-term system performance. There may be no measurable impact on operator attitudes.
None	No problem or negative factor related to system performance is noted.

USABILITY CHECKLIST

Human Factors Principle	Deficiency Rating	Comments
DATA DISPLAY		
1. Ensure that whatever data a user needs for any transaction will be available for display.		
2. Do not overload displays with extraneous data.		
3. For any particular type of data display, maintain consistent format from one display to another.		
4. Ensure that each data display will provide needed context, recapitulating prior data as necessary so that a user does not have to rely on memory to interpret new data.		
5. The wording of displayed data and labels should incorporate familiar terms and the task-oriented jargon of the users.		
6. Choose words carefully and then use them consistently.		
7. When abbreviations are used, choose those abbreviations that are commonly recognized and do not abbreviate words that produce uncommon or ambiguous abbreviations.		
8. Ensure that abbreviations are distinctive so that abbreviations for different words are distinguishable.		
9. When a critical passage merits emphasis to set it apart from other text, highlight that passage by bolding, brightening, color coding, or some auxiliary annotation.		
10. Organize data in some recognizable order to facilitate scanning and assimilation.		
11. In designing text displays, especially text composed for user guidance, strive for simplicity and clarity of wording.		
12. Use consistent logic in the design of graphic displays and maintain standard format, labeling, etc.		
13. Tailor graphic displays to user needs and provide only those data necessary for user tasks.		
14. When graphics contain outstanding or discrepant features that merit attention by a user, consider displaying supplementary text to emphasize that feature.		

Human Factors Principle	Deficiency Rating	Comments
15. When a user's attention must be directed to a portion of a display showing critical or abnormal data, highlight that feature with some distinctive means of coding.		
16. Adopt a consistent organization for the location of various display features from one display to another.		
17. Assign consistent meanings to symbols and other codes from one display to another.		
18. Choose colors for coding based on conventional associations with particular colors.		

Human Factors Principle	Deficiency Rating	Comments
SEQUENCE CONTROL		
1. Defer computer processing until an explicit user action has been taken.		
2. Employ similar means to accomplish ends that are similar, from one transaction to the next, from one task to another, throughout the user interface.		
3. Display some continuous indication of current context for reference by the user.		
4. Adopt consistent terminology for online guidance and other messages to users.		
5. Choose names that are semantically congruent with natural usage, especially for paired opposites (e.g., UP / DOWN).		
6. Ensure that the computer acknowledges every entry immediately; for every action by the user there should be some apparent reaction from the computer.		
7. When a user is performing an operation on some selected display item, highlight that item.		
8. Design the interface software to deal appropriately with all possible control entries, correct and incorrect.		
9. When a user completes correction of an error, require the user to take an explicit action to reenter the corrected material; use the same action for reentry that was used for the original entry.		
10. When a control entry will cause any extensive change in stored data, procedures, and/or system operation, and particularly if that change cannot be easily reversed, notify the user and require confirmation of the action before implementing it. Provide a prompt to confirm actions that will lead to possible data loss.		
USER GUIDANCE		
1. When the computer detects an entry error, display an error message to the user stating what is wrong and what can be done about it.		
2. Make the wording of error messages as specific as possible.		

Human Factors Principle	Deficiency Rating	Comments
3. Make error messages brief but informative.		
4. Adopt neutral wording for error messages; do not imply blame to the user or personalize the computer, or attempt to make a message humorous.		
5. The computer should display an error message only after a user has completed an entry.		
6. Provide reference material describing system capabilities and procedures available to users for online display.		
7. In addition to explicit and implicit aids, permit users to obtain further online guidance by requesting HELP.		

Human Factors Principle	Deficiency Rating	Comments
DATA TRANSMISSION		
1. Choose functional wording for terms used in data transmission, including messages, for initiating and controlling message transmission and other forms of data transfer, and for receiving messages.		
2. Design the data transmission procedures to minimize memory load on the user.		
3. Design the data transmission procedures to minimize required user actions.		
DATA PROTECTION		
1. Provide automatic measures to minimize data loss from computer failure.		
2. Protect data from inadvertent loss caused by the actions of other users.		
3. Provide clear and consistent procedures for different types of transactions, particularly those involving data entry, change and deletion, and error correction.		
4. Ensure that the ease of user actions will match desired ends; make frequent or urgent actions easy to take, but make potentially destructive actions sufficiently difficult that they will require extra user attention.		
5. When displayed data are classified for security purposes, include a prominent indication of security classification in each display.		
6. When a user requests LOG OFF, check pending transactions involving data entry/change and, if data loss seems probable, display an appropriate advisory message to the user.		

Human Factors Principle	Deficiency Rating	Comments
VISUAL DISPLAYS		
1. Sufficient contrast shall be provided between displayed information and the display background to ensure that the required information can be perceived by the operator under all expected lighting conditions.		
2. Displays shall be located and designed so that they may be read to the degree of accuracy required by personnel in normal operating or servicing positions without requiring the operator to assume an uncomfortable, awkward, or unsafe position.		
3. Where alphanumeric characters appear on CRT-like displays, the font style shall allow discrimination of similar characters, such as the letter L and the number 1 and the letter z and the number 2.		

APPENDIX B
CONSENT FORMS, PERSONAL INFORMATION FORMS, AND SURVEYS

INFORMED CONSENT

I, _____, have received a briefing by the FAA about the purpose of the test and have been provided with the opportunity to ask questions. The test will require training on the L³ 3DXTM 6000, followed by screening of checked passenger baggage, some of which will contain simulated explosives. I have also been informed of the safety precautions that will be taken to ensure all testing is to be carried out in a safe and secure manner. Training and testing will require approximately two days.

As part of the data analysis, my data will be combined with that of other individuals and I will no longer be identifiable as a participant. I have been informed that my name will remain CONFIDENTIAL.

I have been informed that I have the right to withdraw from the test, and that the Test Director may terminate my participation in the interest of safety and the test. I also certify that I am at least 18 years of age.

I have been informed that if I have any further concerns or questions, I may contact the Test Director, Brenda Klock, (609) 485-7053; or Dr. J. L. Fobes (609) 485-4944.

Signed: _____

Date: _____ / _____ / _____

Witness: _____

Date: _____ / _____ / _____

PERSONAL INFORMATION FORM

Please provide some basic information about yourself. All of this information will remain confidential.

Name: _____

Date: ____ / ____ / ____

1. At what airport do you work? _____

2. For what screening company do you work? _____

3. What is your gender (please circle your answer)?

Male

Female

4. Have you previously been a certified screener (please circle your answer)?

Yes

No

5. Is English your native or first language (please circle your answer)?

Yes

No

6. Do you wear eyeglasses or corrective lenses (please circle your answer)?

Yes

No

7. Please circle the highest education level that you have completed.

8th Grade

12th Grade

GED

Some College

College Graduate

8. How much computer experience do you have (please circle your answer)?

None

Very Little

Some

A Lot

9. Please circle your age:

18-21

22-29

30-39

40-49

50-59

60-69

70 +

SCREENER SURVEYS ON THE 3DX™ 6000 USABILITY & TRAINING

SUBJECT # _____

DATE _____

The FAA wants to know how useful the 3DX™ 6000 and the training for it is and how they can be made better. Your opinion will be important in improving the 3DX™ 6000 program. Your responses will be kept anonymous, so please give ratings that are your honest opinion of the 3DX™ 6000. We thank you for your help.

If you have any questions while taking these surveys, please ask the FAA representative for help.

SCREENER SURVEY ON THE 3DX™ 6000 USABILITY& TRAINING

SUBJECT # _____

DATE _____

The numbered sentences refer to the machine for which you have just received training. Read each sentence and decide whether you agree with it. Use the rating scale to express your agreement or disagreement with each statement. The scale goes from 1 to 5. If you circle 5, it means you agree very much with the statement. If you circle 1, it means you disagree very much with the statement. Use numbers in the middle to express agreement/disagreement between the extremes. Circle one number for each sentence.

	Strongly Disagree	Disagree	Not Sure	Agree	Strongly Agree
The 3DX™ 6000 improved my ability to find threats in checked bags.	1	2	3	4	5
The 3DX™ 6000 helped me to do my job.	1	2	3	4	5
The purpose of the 3DX™ 6000 was explained very well.	1	2	3	4	5
The training on the use of the 3DX™ 6000 was complete.	1	2	3	4	5
The 3DX™ 6000 slowed down the flow of baggage.	1	2	3	4	5
The 3DX™ 6000 was easy to use.	1	2	3	4	5
The images produced by the 3DX™ 6000 were easy to see.	1	2	3	4	5
I understood and used all of the functions of the 3DX™ 6000.	1	2	3	4	5
The 3DX™ 6000 reminded me of the need to stay aware of possible threats.	1	2	3	4	5
The 3DX™ 6000 helped to teach me to recognize explosive devices.	1	2	3	4	5
I was very tired at the end of the shift because of the 3DX™ 6000.	1	2	3	4	5
The 3DX™ 6000 made my job a lot easier to do.	1	2	3	4	5
The 3DX™ 6000 increased my interest in and enjoyment of the job.	1	2	3	4	5

SCREENER SURVEY ON THE 3DX™ 6000 USABILITY & TRAINING

SUBJECT # _____

DATE _____

II. Please list the most important positive and negative ways in which the 3DX™ 6000 will affect your job.

1. _____

2. _____

3. _____

APPENDIX C

SIGNAL DETECTION THEORY & MEASURES OF SENSITIVITY

SIGNAL DETECTION THEORY

The Signal Detection Theory (SDT) Paradigm.

SDT is based on the concept that the judgment by a person as to what is seen is affected by the expectations of the observer and the payoffs associated with the outcome of judgments to be made. This is the case even in a situation in which the physical input to the receptors is constant (Foley & Morey, 1987). In other words, perception is determined by the interaction of the physical parameters of the stimulus with the subjective control of the perceptual mechanisms by the observer. The theory of signal detection is applicable to situations in which there are two discrete states (e.g., signal and noise) that cannot be easily discriminated (Wickens, 1992).

The Concept of Noise.

A central concept in SDT is that, in any situation, there is noise that can interfere with the detection of a signal. This noise can be generated externally from the person (e.g., noises in a factory other than a warning buzzer) and internally within the person (e.g., miscellaneous neural activity). This noise varies over time, thus forming a distribution of intensity from high to low. The shape of this distribution is assumed to be normal (i.e., bell shaped). When a 'signal' occurs, its intensity is added to that of the background noise. At any given time, a person needs to decide if the sensory input (i.e., what the person senses) consists of only noise or noise plus the signal (Sanders & McCormick, 1987).

Possible Outcomes.

The IED detection tests assess human operators engaged in tasks to detect an environmental event or signal (i.e., a threat object). Based on SDT, there are two responses that represent the screener's detection performance:

YES (a threat object was present), or NO (a threat object was not present).

There are also two signal presentation states:

SIGNAL (a threat object was present), or NOISE (a threat object was not present).

The combination of the screener responses and the signal state produces a 2 x 2 matrix (refer to Figure C-1) comprised of four quadrants. These quadrants are labeled hits, misses, false alarms, and correct rejections.

The Concept of Response Criterion.

One of the major contributions of SDT to the understanding of the detection process was the recognition that people set a criterion along the hypothetical continuum of sensory activity and that this activity is the basis upon which a person makes a decision. The position of the criterion along the continuum determines the probabilities of the four outcomes listed above. The best way to conceptualize this is in Figure C-2. Shown are the hypothetical distributions of sensory activity when only noise is present and when a signal is added to the noise. Notice that the two distributions overlap. That is, sometimes conditions are such that the level of noise will exceed the level of noise plus signal.

As indicated by Wickens (1992), the SDT paradigm assumes that operators perform two stages of information processing in all detection tasks: (1) sensory evidence is aggregated concerning the presence or absence of the signal, and (2) a decision is made about whether this evidence constitutes a signal. According to SDT, external stimuli generate neural activity in the brain. On average, there will be more sensory or neural activity when a signal is present than when it is absent. This neural activity represents the rate of neurons firing in the brain. The response rate for detecting the neural evidence increases with the magnitude of the stimulus (signal) intensity. If there is enough sensory activity, the level of sensory activity exceeds a critical threshold, X , and the operator decides “yes.” If there is too little, the operator decides “no.” This is illustrated in Figure C-2.

SDT postulates that a person sets the criterion level such that whenever the level of sensory activity exceeds that criterion, the person will say there is a signal present. When the activity level is below the criterion, the person will say there is no signal. Figure C-2 also shows four areas corresponding to hits, false alarms, misses, and correct rejections based on the criteria shown in the figure.

Related to the position of the criterion is the operator response criterion quantity (i.e., β). Numerically, the operator response criterion is the ratio (signal to noise) of the height of the two curves in Figure C-2 at a given criterion point. As the criterion is shifted to the right, the response criterion value increases and the person will say signal less often and hence will have fewer hits, but also fewer false alarms. This type of response criterion is considered to be conservative. On the other hand, as the criterion is shifted to the left, β decreases and the person will say signal more often and hence will have more hits, but also more false alarms (Sanders & McCormick, 1987).

For the IED detection tests used in the demonstration testing:

- a. A Hit will be recorded when a screener correctly detects a threat object (e.g., IED) in the bag image.
- b. A False Alarm will be recorded when a screener reports a threat object (e.g., an IED) in the bag image and no threat object is present.
- c. A Miss will be recorded when a screener reports no threat object in the bag image when one actually is present.

- d. A Correct Rejection will be recorded when a screener reports no threat object in the bag image and no threat object is present.

		Signal Presentation State	
		SIGNAL	NOISE
Screener Response	YES	HIT	FALSE ALARM
	NO	MISS	CORRECT REJECTION

FIGURE C-1. MATRIX OF SCREENER RESPONSE AND SIGNAL PRESENTATION STATE

Figure C-2, below, illustrates the key concepts of signal detection theory. Shown are the two hypothetical distributions of internal sensory activity, one generated by noise alone (left side) and the other generated by signal plus noise (right side). The probabilities of the four possible outcomes are depicted as the respective areas under these curves based

$$\beta = b/a$$

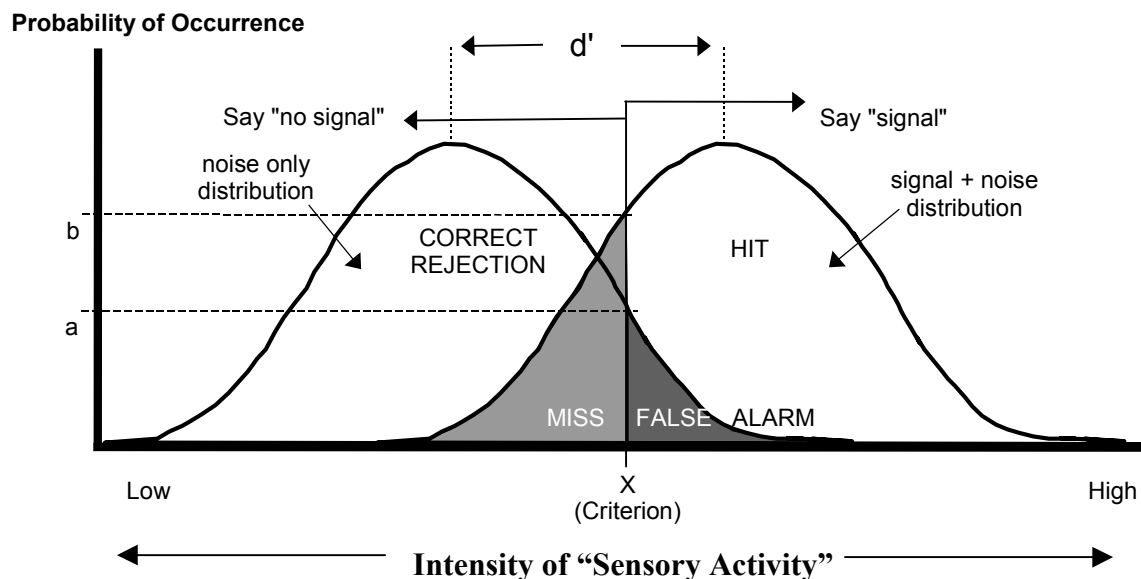


FIGURE C-2. KEY CONCEPTS OF SIGNAL DETECTION THEORY

on the setting of a criterion at X . Here, d' is a measure of sensitivity, and β is a measure of response bias. The letters a and b correspond to the height of the signal-plus-noise and noise-only distributions at the criterion.

Influencing Response Criterion.

SDT postulates two major variables that influence the setting of the criterion: (1) the likelihood of observing a signal, and (2) the costs and benefits associated with the four possible outcomes. Consider the first. If field intelligence indicated that a terrorist had threatened a flight, thus increasing the likelihood of a threat object in a carry-on bag, the screener would be more likely to call an object on the X-ray a threat object. That is, as the probability of a signal increases, the response criterion is lowered (reducing β), so that anything remotely suggesting a signal (a threat object in our example) might be called a signal (Sanders & McCormick, 1987).

With regard to the second factor, costs and benefits, again consider the tasks of the airport screener. What is the cost of a false alarm (saying there may be a threat object when there is not)? The bag is hand searched and throughput at the checkpoint is decreased. What is the cost of a miss (saying that not threat object exists when there is one)? The threat object may get on to the aircraft and an act of terrorism may occur. Under these circumstances, the screener may set a low criterion and be more willing to call a suspicious object a potential threat. But what if the bag has to pass through three different checkpoints with different types of sophisticated detection equipment? In this case, the screener would set a more conservative criterion and would be less likely to call it a threat object.

Concept of Sensitivity.

In addition to the concept of response criterion, SDT also includes a measure of the person's sensitivity. That is, the keenness or resolution of the sensory system. Further, SDT postulates that the response criterion and sensitivity are independent of each other. In SDT terms, sensitivity is measured by the degree of separation between the two distributions shown in Figure A-2. The sensitivity measure is called d' and corresponds to the separation of the two distributions expressed in units of standard deviation (it is assumed that the standard deviations of the two distributions are equal). The greater the separation, the greater the sensitivity and the greater the d' . In most applications of SDT, d' ranges between 0.5 and 2.0.

Some signal generation systems may create more noise than others, and some people may have more internal noise than others. The greater the amount of noise, the smaller d' will be. Also, the weaker and less distinct the signal, the smaller d' will be. Another factor that influences d' is the ability of the person to remember the physical characteristics of a signal. For example, when memory aids are supplied, sensitivity increase (Wickens, 1992). In the case of the X-Ray screener, better X-ray equipment and images, or X-rays

of actual threat objects to compare with the X-ray under consideration, should increase sensitivity.

Procedures to Calculate SDT Probabilities.

- a. In SDT, the detection values are expressed as probabilities.
- b. The probability of a hit (P_h), miss (P_m), false alarm (P_{fa}), and correct rejection (P_{cr}) are determined by dividing the number of occurrences in a cell (refer to Figure A-1) by the total number of occurrences in a column.
- c. The P_h (also referred to as the probability of detection, P_d) will be calculated by dividing the number of threats detected (number of hits) by the total number of hits and misses:
$$P_m = 1 - P_d$$
- d. The P_{fa} will be determined by the number of false alarms divided by the total number of false alarms and correct rejections:
$$P_{cr} = 1 - P_{fa}$$

Procedures to Calculate the Response Criterion Measure β .

- a. Find the false alarm rate from the outcome matrix in the HIT/FA column of Table C-1.
- b. Read across the table to the ORD column (for ordinate, the height of the curve).

TABLE C-1. REPRESENTATIVE Z-SCORES AND ORDINATE VALUES OF THE
NORMAL CURVE FOR DIFFERENT RESPONSE PROBABILITIES
TO CALCULATE c AND d'

HIT/FA	Z	ORD	HIT/FA	Z	ORD
.01	2.33	0.03	.50	0.00	0.40
.02	2.05	0.05	.55	-0.12	0.40
.03	1.88	0.07	.60	-0.25	0.39
.04	1.75	0.09	.65	-0.38	0.37
.05	1.64	0.10	.70	-0.52	0.35
.08	1.40	0.15	.75	-0.67	0.32
.10	1.28	0.18	.80	-0.84	0.28
.13	1.13	0.21	.82	-0.92	0.26
.15	1.04	0.23	.85	-1.04	0.23
.18	0.92	0.26	.88	-1.18	0.20
.20	0.84	0.28	.90	-1.28	0.18
.25	0.67	0.32	.92	-1.40	0.15
.30	0.52	0.35	.95	-1.64	0.10
.35	0.38	0.37	.96	-1.75	0.09
.40	0.25	0.39	.97	-1.88	0.07
.45	0.12	0.40	.98	-2.05	0.05
.50	0.00	0.40	.99	-2.33	0.03